Contribution ID: **73**   Type: **Oral**

Managing Experiment Data with Ease at the Advanced Photon Source

Tuesday 20 September 2022 14:50 (20 minutes)

Managing Experiment Data with Ease at the Advanced Photon Source

Hannah Parraga, Sinisa Veseli, John Hammonds, Steven Henke, Nicholas Schwarz

Advanced Photon Source, Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439, USA

Data are essential to the scientific discoveries enabled by experiments performed at the Advanced Photon Source (APS). At present, the APS generates approximately 10 PB of raw experimental data per year from its sixty-eight operating beamlines that house over 100 unique instruments. This data is generated as a part of over 6,000 annual experiments performed by over 5,500 facility users each year. Similar to other synchrotrons, the amount of data generated at the APS continues to quickly increase due to beamline advances, such as new measurement techniques, technological advances in detectors and instrumentation, multi-modal instruments that can acquire several measurements in a single experiment, and advanced data processing algorithms. This trend is expected to continue in the future.

As a scientific user facility, the APS presents several unique challenges for data management. Beamlines can perform multiple experiment techniques, use different types of detectors, produce data at different rates, use multiple data formats, use machines with different operating systems, and execute various processing work-flows. Additionally, the users themselves vary. They come from different research institutions, universities, and industries, but all must be able to access their data after leaving the facility. They may want their data immediately or several years after it is created. They may be conducting experiments independently and remotely, or in person with close involvement with beamline staff. Also, beamline scientists have different levels of technical expertise. Some desire a hands-off approach to data management and some want the flexibility to program custom tools. The APS must have a data management solution that addresses these unique challenges.

This presentation covers the features of the APS Data Management System, which provides tools that beamline staff can use to support their users. A command line interface and graphical user interface give users the ability to upload data to long-term storage. The built-in workflow engine allows data to be processed with any given set of shell commands. Using Globus, data is secure but also accessible to users from their home institution. Data is secured on local beamline machines with tools for managing file system permissions. Furthermore, users are able to catalog metadata to include additional information about the experiments alongside their results.

Although the APS Data Management System addresses many needs, further development is underway of additional features to provide users with an improved data management experience. This includes streaming data directly from detectors to storage to decrease the transfer time as data rates and volumes continue to increase. Workflows are being developed which publish to common data portals for visualizing results. Interfacing with the tape archives of the Argonne Leadership Computing Facility will allow more data to be stored for longer periods.

Email address of presenting author

hparraga@anl.gov

Author: PARRAGA, Hannah

Co-authors: HAMMONDS, John (Argonne National Laboratory); SCHWARZ, Nicholas (Argonne National Laboratory); VESELI, Sinisa (Argonne National Laboratory); HENKE, Steven (Argonne National Laboratory)

Presenter: PARRAGA, Hannah

Track Classification: NOBUGS 2022